数理解析研究所講究録 1268 巻 2002 年 1-6

n-Insertion on Languages

京都産業大学・理学部　伊藤 正美
京都産業大学・大学院　杉浦 亮

Masami Ito [1], Ryo Sugiura [2]

Department of Mathematics [1], Graduate School [2]

Kyoto Sangyo University [1, 2], Kyoto 603-8555, Japan

Abstract

In this paper, we define the n-insertion $A \rhd^{[n]} B$ of a language A into a language B and provide some properties of n-insertions. For instance, the n-insertion of a regular language into a regular language is regular but the n-insertion of a context-free language into a context-free language is not always context-free. However, it can be shown that the n-insertion of a regular (context-free) language into a context-free (regular) language is context-free. We also consider the decomposition of regular languages under n-insertion.

1 Introduction

Let $u, v \in X^*$ and let n be a positive integer. Then the *n-insertion* of u into v, i.e. $u \rhd^{[n]} v$, is defined as $\{v_1u_1v_2u_2 \dots v_nu_nv_{n+1} \mid u = u_1u_2 \dots u_n, u_1, u_2, \dots, u_n \in X^*, v = v_1v_2 \dots v_nv_{n+1}, v_1, v_2, \dots, v_n, v_{n+1} \in X^*\}$. For languages $A, B \subseteq X^*$, the *n-insertion* $A \rhd^{[n]} B$ of A into B is defined as $\bigcup_{u\in A, v\in B} u \rhd^{[n]} v$. The shuffle product $A \diamond B$ of A and B is defined as $\bigcup_{n\geq 1} A \rhd^{[n]} B$. In Section 2, we provide some properties of n-insertions. For instance, the n-insertion of a regular language into a regular language is regular but the n-insertion of a context-free language into a context-free language is not always context-free. However, it can be shown that the n-insertion of a regular (context-free) language into a context-free (regular) language is context-free. In Section 3, we prove that, for a given regular language $L \subseteq X^*$ and a positive integr n, it is decidable whether $L = A \rhd^{[n]} B$ for some nontrivial regular languages $A, B \subseteq X^*$. Here a language $C \subseteq X^*$ is said to be *nontrivial* if $C \neq \{\epsilon\}$ where ϵ is the empty word. Regarding definitions and notations concerning formal languages and automata, not defined in this paper, refer, for instance,

2 Shuffle Product and n-Insertion

First, we consider the shuffle product of languages.

Lemma 1 *Let $A, B \subseteq X^*$ be regular languages. Then $A \diamond B$ is a regular language.*

Proof By $\overline{X}$ we denote the new alphabet $\{\overline{a} \mid a \in X\}$. Let $\mathcal{A} = (S, X, \delta, s_0, F)$ be a finite deterministic automaton with $\mathcal{L}(\mathcal{A}) = A$ and let $\mathcal{B} = (T, X, \theta, t_0, G)$ be a finite deterministic automaton with $\mathcal{L}(\mathcal{B}) = B$. Define the automaton $\overline{\mathcal{B}} = (T, \overline{X}, \overline{\theta}, t_0, G)$ as $\overline{\theta}(t, \overline{a}) = \theta(t, a)$ for any $t \in T$ and $a \in X$. Let ρ be the homomorphism of $(X \cup \overline{X})^*$ onto X^* defined as $\rho(a) = \rho(\overline{a}) = a$ for any $a \in X$. Moreover, let $\mathcal{L}(\overline{\mathcal{B}}) = \overline{B}$. Then $\rho(\overline{B}) = \{\rho(\overline{u}) \mid \overline{u} \in \overline{B}\} = B$ and $\rho(A \diamond \overline{B}) = A \diamond B$. Hence, to prove the lemma, it is enough to show that $A \diamond \overline{B}$ is a regular language over $X \cup \overline{X}$. Consider the automaton $\mathcal{A} \diamond \overline{\mathcal{B}} = (S \times T, X \cup \overline{X}, \delta \diamond \overline{\theta}, (s_0, t_0), F \times G)$ where $\delta \diamond \overline{\theta}((s, t), a) = (\delta(s, a), t)$ and $\delta \diamond \overline{\theta}((s, t), \overline{a}) = (s, \theta(t, a))$ for any $(s, t) \in S \times T$ and $a \in X$. Then it is easy to see that $w \in \mathcal{L}(\mathcal{A} \diamond \overline{\mathcal{B}})$ if and only if $w \in A \diamond \overline{B}$, i.e. $A \diamond \overline{B}$ is regular. This completes the proof of the lemma.

Proposition 2 *Let $A, B \subseteq X^*$ be regular languages and let n be a positive integer. Then $A \rhd^{[n]} B$ is a regular language.*

Proof Let the notations of $\overline{X}$, $\overline{B}$ and ρ be the same as above. Notice that $A \rhd^{[n]} \overline{B} = (A \diamond \overline{B}) \cap (\overline{X}^* X^*)^n \overline{X}^*$. Since $(\overline{X}^* X^*)^n \overline{X}^*$ is regular, $A \rhd^{[n]} \overline{B}$ is regular. Consequently, $A \rhd^{[n]} B = \rho(A \rhd^{[n]} \overline{B})$ is regular.

Remark 3 The n-insertion of a context-free language into a context-free language is not always context-free. For instance, it is well known that $A = \{a^n b^n \mid n \geq 1\}$ and $B = \{c^n d^n \mid n \geq 1\}$ are context-free languages over $\{a, b\}$ and $\{c, d\}$, respectively. However, since $(A \rhd^{[2]} B) \cap a^+ c^+ b^+ d^+ = \{a^n c^m b^n d^m \mid n, m \geq 1\}$ is not context-free, $A \rhd^{[2]} B$ is not context-free.

Now consider the n-insertion of a regular (context-free) language into a context-free (regular) language.

Lemma 4 *Let $A \subseteq X^*$ be a regular language and let $B \subseteq X^*$ be a context-free language. Then $A \diamond B$ is a context-free language.*

Proof The notations which we will use for the proof are assumed to be the same as above. Let $\mathcal{A} = (S, X, \delta, s_0, F)$ be a finite deterministic automaton with $\mathcal{L}(\mathcal{A}) = A$ and let $\mathcal{B} = (T, X, \Gamma, \theta, t_0, \epsilon)$ be a pushdown automaton with $\mathcal{N}(\mathcal{B}) = B$. Let $\overline{\mathcal{B}} = (T, \overline{X}, \Gamma, \overline{\theta}, t_0, \gamma_0, \epsilon)$ be a pushdown automaton such that $\overline{\theta}(t, \overline{a}, \gamma) = \theta(t, a, \gamma)$ for any $t \in T, a \in X \cup \{\epsilon\}$ and $\gamma \in \Gamma$. Then $\rho(\mathcal{N}(\overline{\mathcal{B}})) = B$. Now define the pushdown automaton $\mathcal{A} \diamond \overline{\mathcal{B}} = (S \times T, X \cup \overline{X}, \Gamma \cup \{\#\}, \delta \diamond \overline{\theta}, (s_0, t_0), \gamma_0, \epsilon)$ as follows: (1) $\forall a \in X, \delta \diamond \overline{\theta}((s_0, t_0), a, \gamma_0) = \{((\delta(s_0, a), t_0), \#\gamma_0)\}$, $\delta \diamond \overline{\theta}((s_0, t_0), \overline{a}, \gamma_0) = \{((s_0, t'), \#\gamma') \mid (t', \gamma') \in \overline{\theta}(t_0, \overline{a}, \gamma_0)\}$. (2) $\forall a \in X, \forall (s,t) \in S \times T, \forall \gamma \in \Gamma \cup \{\#\}, \delta \diamond \overline{\theta}((s,t), a, \gamma) = \{((\delta(s,a), t), \gamma)\}$. (3) $\forall a \in X, \forall (s,t) \in S \times T, \forall \gamma \in \Gamma, \delta \diamond \overline{\theta}((s,t), \overline{a}, \gamma) = \{((s, t'), \gamma') \mid (t', \gamma') \in \overline{\theta}(t, \overline{a}, \gamma)\}$. (4) $\forall (s,t) \in F \times T, \delta \diamond \overline{\theta}((s,t), \epsilon, \#) = \{((s,t), \epsilon)\}$.

Let $w = \overline{v}_1 u_1 \overline{v}_2 u_2 \ldots \overline{v}_n u_n \overline{v}_{n+1}$ where $u_1, u_2, \ldots, u_n \in X^*$ and $\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_{n+1} \in \overline{X}^*$. Assume $\delta \diamond \overline{\theta}((s_0, t_0), w, \gamma_0) \neq \emptyset$. Then we have the following configuration: $((s_0, t_0), w, \gamma_0) \vdash^*_{\mathcal{A}\diamond\overline{\mathcal{B}}} ((\delta(s_0, u_1 u_2 \ldots u_n), t'), \epsilon, \# \ldots \#\gamma')$ where $(t', \gamma') \in \overline{\theta}(t_0, \overline{v}_1 \overline{v}_2 \ldots \overline{v}_{n+1}, \gamma_0)$. If $w \in \mathcal{N}(\mathcal{A}\diamond\overline{\mathcal{B}})$, then $(\delta(s_0, u_1 u_2 \ldots u_n), t'), \epsilon, \# \ldots \#\gamma') \vdash^*_{\mathcal{A}\diamond\overline{\mathcal{B}}} (\delta(s_0, u_1 u_2 \ldots u_n), t'), \epsilon, \epsilon)$. Therefore, $(\delta(s_0, u_1 u_2 \ldots u_n), t') \in F \times T$ and $\gamma' = \epsilon$. This means that $u_1 u_2 \ldots u_n \in A$ and $\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_{n+1} \in \overline{B}$. Hence $w \in A \times \overline{B}$. Now let $w \in A \times \overline{B}$. Then, by the above configuration, we have $((s_0, t_0), w, \gamma_0) \vdash^*_{\mathcal{A}\diamond\overline{\mathcal{B}}} ((\delta(s_0, u_1 u_2 \ldots u_n), t'), \epsilon, \# \ldots \#) \vdash^*_{\mathcal{A}\diamond\overline{\mathcal{B}}} ((\delta(s_0, u_1 u_2 \ldots u_n), t'), \epsilon, \epsilon)$ and $w \in \mathcal{N}(\mathcal{A} \diamond \overline{\mathcal{B}})$. Thus $A \diamond \overline{B} = \mathcal{N}(\mathcal{A} \diamond \overline{\mathcal{B}})$ and $A \diamond \overline{B}$ is context-free. Since $\rho(A \diamond \overline{B}) = A \diamond B$, $A \diamond B$ is context-free.

Proposition 5 *Let $A \subseteq X^*$ be a regular (context-free) language and let $B \subseteq X^*$ be a context-free (regular) language. Then $A \rhd^{[n]} B$ is a context-free language.*

Proof We consider the case that $A \subseteq X^*$ is regular and $B \subseteq X^*$ is context-free. Since $A \rhd^{[n]} \overline{B} = (A \diamond \overline{B}) \cap (\overline{X}^* X^*)^n \overline{X}^*$ and $(\overline{X}^* X^*)^n \overline{X}^*$ is regular, $A \rhd^{[n]} \overline{B}$ is context-free. Consequently, $A \rhd^{[n]} B = \rho(A \rhd^{[n]} \overline{B})$ is context-free.

3 Decomposition

Let $L \subseteq X^*$ be a regular language and let $\mathcal{A} = (S, X, \delta, s_0, F)$ be a finite deterministic automaton accepting the language L, i.e. $\mathcal{L}(\mathcal{A}) = L$. For $u, v \in X^*$, by $u \sim v$ we denote the equivalence relation of finite index on X^* such that $\delta(s, u) = \delta(s, v)$ for any $s \in S$. Then it is well known that for any $x, y \in X^*$, $xuy \in L \Leftrightarrow xvy \in L$ if $u \sim v$. Let $[u] = \{v \in X^* \mid u \sim v\}$ for $u \in X^*$. It is easy to see that $[u]$ can be effectively constructed using $\mathcal{A}$ for any $u \in X^*$. Now let n be a positive integer. We consider the decomposition $L = A \rhd^{[n]} B$. Let $K_n = \{([u_1],[u_2],\ldots,[u_n]) \mid u_1,u_2,\ldots,u_n \in X^*\}$. Notice that K_n is a finite set.

Lemma 6 *There is an algorithm to construct K_n.*

Proof Obvious from the fact that $[u]$ can be effectively constructed for any $u \in X^*$ and $\{[u] \mid u \in X^*\} = \{[u] \mid u \in X^*, |u| \leq |S|^{|S|}\}$. Here $|u|$ and $|S|$ denote the length of u and the cardinality of S, respectively.

For $u \in X^*$, we define $\rho_n(u)$ by $\{([u_1],[u_2],\ldots,[u_n]) \mid u = u_1u_2\ldots u_n, u_1, u_2, \ldots, u_n \in X^*\}$. Let $\mu = ([u_1],[u_2],\ldots,[u_n]) \in K_n$ and let $B_\mu = \{v \in X^* \mid \forall v = v_1v_2\ldots v_nv_{n+1}, v_1, v_2, \ldots .v_n, v_{n+1} \in X^*, \{v_1\}[u_1]\{v_2\}[u_2]\ldots\{v_n\}[u_n] \{v_{n+1}\} \subseteq L\}$.

Lemma 7 *$B_\mu \subseteq X^*$ is a regular language and it can be effectively constructed.*

Proof Let $S^{(i)} = \{s^{(i)} \mid s \in S\}, 0 \leq i \leq n$, and let $\tilde{S} = \bigcup_{0\leq i\leq n} S^{(i)}$. We define the following nondeterministic automaton $\tilde{\mathcal{A}}' = (\tilde{S}, X, \tilde{\delta}, \{s_0^{(0)}\}, S^{(n)} \setminus F^{(n)})$ with ϵ-move where $F^{(n)} = \{s^{(n)} \mid s \in F\}$. The state transition relation $\tilde{\delta}$ is defined as follows:

$\tilde{\delta}(s^{(i)}, a) = \{\delta(s,a)^{(i)}, \delta(s, au_{i+1})^{(i+1)}\}$ for any $a \in X \cup \{\epsilon\}$ and $i = 0, 1, \ldots, n-1$ and $\tilde{\delta}(s^{(n)}, a) = \{\delta(s,a)^{(n)}\}$ for any $a \in X$.

Let $v \in \mathcal{L}(\tilde{\mathcal{A}}')$. Then $\delta(s_0, v_1u_1v_2u_2\ldots v_nu_nv_{n+1})^{(n)} \in \tilde{\delta}(s_0^{(0)}, v_1v_2\ldots v_nv_{n+1}) \cap (S^{(n)} \setminus F^{(n)})$ for some $v = v_1v_2\ldots v_nv_{n+1}, v_1, v_2, \ldots, v_n, v_{n+1} \in X^*$. Hence $v_1u_1v_2u_2\ldots v_nu_nv_{n+1} \notin L$, i.e. $v \in X^* \setminus B_\mu$. Now let $v \in X^* \setminus B_\mu$. Then there exists $v = v_1v_2\ldots v_nv_{n+1}, v_1, v_2, \ldots, v_n.v_{n+1} \in X^*$ such that $v_1u_1v_2u_1\ldots v_nu_nv_{n+1}$ $\notin L$. Therefore, $\tilde{\delta}(s_0^{(0)}, v_1v_2\ldots v_nv_{n+1}) \in S^{(n)} \setminus F^{(n)}$, i.e. $v = v_1v_2\ldots v_nv_{n+1} \in \mathcal{L}(\tilde{\mathcal{A}}')$. Consequently, $B_\mu = X^* \setminus \mathcal{L}(\tilde{\mathcal{A}}')$ and B_μ is regular. Notice that $X^* \setminus \mathcal{L}(\tilde{\mathcal{A}}')$ can be effectively constructed.

Symmetrically, consider $\nu = ([v_1],[v_2],\ldots,[v_n],[v_{n+1}]) \in K_{n+1}$ and $A_\nu = \{u \in X^* \mid \forall u = u_1u_2\ldots u_n, u_1, u_2, \ldots, u_n \in X^*, [v_1]\{u_1\}[v_2]\{u_2\}\ldots[v_n]\{u_n\} [v_{n+1}] \subseteq L\}$.

Lemma 8 *$A_\nu \subseteq X^*$ is a regular language and it can be effectively constructed.*

Proof Let $S^{(i)} = \{s^{(i)} \mid s \in S\}, 1 \leq i \leq n+1$, and let $\overline{S} = \bigcup_{1\leq i\leq n+1} S^{(i)}$. We define the following nondeterministic automaton $\overline{\mathcal{B}}' = (\overline{S}, X, \overline{\delta}, \{\delta(s_0, v_1)^{(1)}\},$

$S^{(n+1)} \setminus F^{(n+1)})$ with ϵ-move where $F^{(n+1)} = \{s^{(n+1)} \mid s \in F\}$. The state transition relation $\overline{\delta}$ is defined as follows:

$\overline{\delta}(s^{(i)}, a) = \{\delta(s,a)^{(i)}, \delta(s, au_{i+1})^{(i+1)}\}$ for any $a \in X \cup \{\epsilon\}$ and $i = 1, 2, \ldots, n$.

By the same way as in the proof of Lemma 6, we can prove that $A_\nu = X^* \setminus \mathcal{L}(\overline{\mathcal{B}'})$. Therefore, A_ν is regular. Notice that $X^* \setminus \mathcal{L}(\overline{\mathcal{B}'})$ can be effectively constructed.

Proposition 9 *Let $A, B \subseteq X^*$ and let $L \subseteq X^*$ be a regular language. If $L = A \rhd^{[n]} B$, then there exist regular languages $A', B' \subseteq X^*$ such that $A \subseteq A', B \subseteq B'$ and $L = A' \rhd^{[n]} B'$.*

Proof Put $B' = \bigcap_{\mu \in \rho_n(A)} B_\mu$. Let $v \in B$ and let $\mu \in \rho_n(A)$. Since $\mu \in \rho_n(A)$, there exists $u \in A$ such that $\mu = ([u_1], [u_2], \ldots, [u_n])$ and $u = u_1 u_2 \ldots u_n, u_1, u_2,$
$\ldots, u_n \in X^*$. By $u \rhd^{[n]} v \subseteq L$, we have $\{v_1\}[u_1]\{v_2\}[u_2]\ldots\{v_n\}[u_n]\{v_{n+1}\} \subseteq L$ for any $v = v_1 v_2 \ldots v_n v_{n+1}, v_1, v_2, \ldots, v_n, v_{n+1} \in X^*$. This means that $v \in B_\mu$. Thus $B \subseteq \bigcap_{\mu \in \rho_n(A)} B_\mu = B'$. Now assume that $u \in A$ and $v \in B'$. Let $u = u_1 u_2 \ldots u_n, u_1, u_2, \ldots, u_n \in X^*$ and let $\mu = ([u_1], [u_2], \ldots, [u_n]) \in \rho_n(u) \subseteq \rho_n(A)$. By $v \in B' \subseteq B_\mu$, $v_1 u_1 v_2 u_2 \ldots v_n u_n v_{n+1} \in \{v_1\}[u_1]\{v_2\}[u_2]\ldots$ $\{v_n\}[u_n]\{v_{n+1}\} \subseteq L$ for any $v = v_1 v_2 \ldots v_n v_{n+1}, v_1, v_2, \ldots, v_n, v_{n+1} \in X^*$. Hence $u \rhd^{[n]} v \subseteq L$ and $A \rhd^{[n]} B' \subseteq L$. On the other hand, since $B \subseteq B'$ and $A \rhd^{[n]} B = L$, we have $A \rhd^{[n]} B' = L$. Symmetrically, put $A' = \bigcap_{\nu \in \rho_{n+1}(B')} A_\nu$. By the same way as the above, we can prove that $A \subseteq A'$ and $L = A' \rhd^{[n]} B'$.

Theorem 10 *For any regular language $L \subseteq X^*$ and a positive integer n, it is decidable whether $L = A \rhd^{[n]} B$ for some nontrivial regular languages $A, B \subseteq X^*$.*

Proof Let $\mathbf{A} = \{A_\nu \mid \nu \in K_{n+1}\}$ and $\mathbf{B} = \{B_\mu \mid \mu \in K_n\}$. By the preceding lemmata, $\mathbf{A}$, $\mathbf{B}$ are finite sets of regular languages which can be effectively constructed. Assume that $L = A \rhd^{[n]} B$ for some nontrivial regular languages $A, B \subseteq X^*$. In this case, by Proposition 8, there exist regular languages $A \subseteq A'$ and $B \subseteq B'$ which are an intersection of languages in $\mathbf{A}$ and an intersection of languages in $\mathbf{B}$, respectively. It is obvious that A', B' are nontrivial languages. Thus we have the following algorithm: (1) Take any languages from $\mathbf{A}$ and let A' be their intersection. (2) Take any languages from $\mathbf{B}$ and let B' be their intersection. (3) Calculate $A' \rhd^{[n]} B'$. (4) If

$A' \rhd^{[n]} B' = L$, then the output is "YES". (5) If the output is "NO", search another pair of $\{A', B'\}$ until obtaining the output "YES". (6) This procedure terminates after a finite-step trial. (7) Once we get the output "YES", then $L = A \rhd^{[n]} B$ for some nontrivial regular languages $A, B \subseteq X^*$. (8) Otherwise, there are no such decompositions.

Let n be a positive integer. By $\mathcal{F}(n, X)$, we denote the class of finite languages $\{L \subseteq X^* \mid max\{|u| \mid u \in L\} \leq n\}$. Then the following result by C. Câmpeanu et al. ([1]) can be obtained as a corollary of Theorem 10.

Corollary *For a given positive integer n and a regular language $A \subseteq X^*$, the problem whether $A = B \diamond C$ for a nontrivial language $B \in \mathcal{F}(n, X)$ and a nontrivial regular language $C \subseteq X^*$ is decidable.*

Proof Obvious from the following fact: If $u, v \in X^*$ and $|u| \leq n$, then $u \diamond v = u \rhd^{[n]} v$.

The proof of the above corollary was given by the different way in ([3]) using the following result: *Let $A, L \subseteq X^*$ be regular languages. Then it is decidable whether there exists a regular languages $B \subseteq X^*$ such that $L = A \diamond B$.*

References

[1] C. Câmpeanu, K. Salomaa, S. Vágvölgyi, Shuffle quotient and decompositions, Lecture Notes in Computer Science (Springer), to appear.

[2] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, Reading MA,1979.

[3] M. Ito, Shuffle decomposition of regular languages, Journal of Universal Computer Science 8 (2002), 257 - 259.